

# La Vision 50/50 pour le trafic sur Internet

Version 2.0, janvier 2024

## Introduction

La Vision 50/50 définit la stratégie de l'Internet Society visant à rallier les efforts de plusieurs parties prenantes et les ressources internationales et nationales pour garantir qu'au moins 50 % de l'ensemble du trafic généré localement dans les économies sélectionnées reste local d'ici 2025. La réalisation de cet objectif ambitieux renforcera la qualité d'Internet et réduira les coûts d'accès pour les particuliers.

Une étude de 2012 a établi une base de référence pour les points d'échange Internet (IXP) au Kenya et au Nigéria, et une étude de suivi en 2020 a montré que les niveaux de trafic local sont passés de 30 % à 70 %. Cette étude a permis de mieux comprendre l'impact du peering sur l'infrastructure locale. Il a été constaté que l'augmentation du trafic local a permis aux réseaux participants de réaliser d'importantes économies et a placé ces deux pays dans une position plus forte pour participer à l'économie numérique. Dans le même temps, l'Afrique du Sud a localisé plus de 80 % de son trafic local et bénéficie désormais d'un Internet stable, résilient, de haute qualité et abordable.

L'objectif de l'étude de la Vision 50/50 est de mesurer le pourcentage du trafic Internet qui est servi à partir d'un serveur ou d'un cache national. Ce document présente la méthodologie de l'Internet Society pour atteindre l'objectif susmentionné.

## Définitions

- Trafic local : le trafic qui reste local et ne quitte pas le pays.
- Contenu local : le contenu Internet hébergé dans le pays (par exemple, site Web d'information, services de gouvernement virtuel, etc.) et
- RIPE Atlas : système de mesure d'Internet géré par le RIPE NCC, composé de sondes (points d'observation) et de points d'ancrage (cibles). Les sondes peuvent être utilisées pour effectuer plusieurs mesures (latence, traceroute, DNS, SSL, HTTP) à partir de la périphérie.

- CDN : les réseaux de diffusion des données sont responsables de la diffusion des données vers les périphéries.
- Cache de contenu : équipement d'hébergement de contenu placé par un fournisseur de contenu à proximité des utilisateurs finaux.
- Point d'observation : endroit d'où la mesure est effectuée.
- Réseau périphérique : le réseau où se trouvent les utilisateurs finaux (consommateurs).

## Manque de données sur le volume de trafic

Il n'est pas facile de mesurer les niveaux de trafic local et non local dans un pays. Les fournisseurs d'accès savent généralement quels sont les services les plus utilisés et quels sont ceux qui consomment le plus de bande passante sur leur réseau. Les données relatives aux services les plus utilisés sont rarement publiées. Dans de rares cas et sur la base d'un accord préalable, les FAI fournissent des données de trafic s-flow (échantillons de flux), à partir desquelles l'IP de destination peut être extraite. En l'absence de données de référence, nous utiliserons le pourcentage de sites Web desservis localement (en mode natif ou par le biais d'un cache CDN) en examinant les 1 000 principaux sites Web d'un pays.

## Google CrUX

Nous avons utilisé la liste des 1 000 principaux sites Web du [Google Chrome UX Report](#) (CrUX). Il s'agit d'un ensemble de données publiques et d'une initiative de Google qui vise à fournir des indicateurs de performance concrets sur la façon dont les utilisateurs se servent du Web. Elle rassemble des données anonymisées sur les performances des utilisateurs qui ont accepté de partager leurs données de navigation avec Google, principalement par l'intermédiaire du navigateur Chrome. CrUX produit un rapport mensuel des sites Web les plus utilisés par pays.

## Comment déterminons-nous le niveau de localité des contenus

?

1. Nous devons déterminer quel CDN est utilisé par chaque site Web du Top 1 000. Pour ce faire, nous inspectons le site Web afin d'identifier les éléments suivants :
  - a. En-têtes de serveur HTTPS : les en-têtes tels que « server » ou « via » contiennent généralement des indications géographiques.
  - b. Enregistrements CNAME : un enregistrement de nom canonique (CNAME RECORD) est un type d'enregistrement dans les serveurs de noms de domaine (serveurs DNS) qui crée un alias d'un domaine vers un autre. En inspectant un enregistrement CNAME, nous

- pouvons voir le nom de domaine lié à un CDN, avant de le diriger vers la ressource souhaitée.
- c. Données WHOIS : en utilisant l'adresse IP du domaine, nous pouvons faire une recherche WHOIS pour déduire quels CDN sont utilisés.
  - d. En-tête du serveur : certains en-têtes de réponse HTTP fournissent des indications géographiques.
2. En utilisant des points d'observation locaux à l'intérieur du pays, nous effectuons des mesures actives vers un ensemble sélectionné de cibles (CDN, applications, sites Web) afin d'extraire des indications géographiques et d'autres informations de géolocalisation. À l'aide des informations ci-dessus, nous pouvons indiquer en toute confiance qu'un site Web est local.
  3. Pour les sites Web hébergés nativement, nous utilisons la géolocalisation pour déterminer l'emplacement de l'adresse IP. L'adresse IP est obtenue en exécutant la requête à partir d'un point d'observation local par l'intermédiaire d'un résolveur DNS local.
  4. Nous calculons ensuite le rapport entre les sites Web hébergés localement et les sites Web hébergés à distance.

## Exemple

Voici la marche à suivre pour calculer le pourcentage de sites Web hébergés localement :

1. Nous prenons le Top 1 000 des sites Web de Google CrUX pour le pays souhaité.
2. Nous classons les sites Web par <nom de CDN> s'ils sont hébergés par un CDN ou en indiquant « Native » s'ils sont hébergés localement.
3. Nous utilisons la géolocalisation de l'IP pour déterminer l'emplacement des domaines hébergés nativement.

CDN	Nombre	Lieu
Cloudflare	454	Local
Akamai	92	Externe
Cloudfront	55	Local
Google	51	Local
Fastly	25	Externe
Facebook	8	Local

Yahoo!	6	Externe
Amazon AWS	5	Externe
EdgeCast	4	Externe
Level3	3	Externe
Hébergement natif	250	Local : 100 Externe : 150

Tableau 1 : Répartition de l'utilisation des CDN pour les 1 000 principaux sites Web d'un pays

4. Le résultat est le rapport entre les sites Web locaux et les sites Web externes.

## Résultat

Nous prévoyons de produire un rapport détaillé sur la part du trafic qui reste locale sur la base de la méthodologie ci-dessus d'ici à la fin février 2024. Nous développerons également un tableau de bord sur notre plateforme Pulse pour visualiser les données au deuxième trimestre 2024. Nous diviserons le Top 1 000 des sites Web en différents ensembles de données afin de présenter différentes perspectives sur la localité du trafic : (1) top 1 000 des sites Web (2) top 10 des applications (Netflix, Disney+, etc.) (3) top 100 des sites Web utilisant le ccTLD du pays (4) sites Web gouvernementaux (5) sites Web d'informations locales (6) principaux acteurs CDN.